amazon | science

Compatibility-aware Heterogeneous Visual Search

CVPR VIRTUAL JUNE 19-25

Rahul Duggal[1,2], Hao Zhou[1], Shuo Yang[1], Yuanjun Xiong[1], Wei Xia[1], Zhuowen Tu[1], Stefano Soatto[1]
[1]AWS AI, [2]Georgia Tech

Abstract: We tackle the problem of visual search under resource constraints. Existing systems use the same embedding model to compute embeddings for the query and gallery images necessitating a hard accuracy-efficiency tradeoff. We mitigate the tradeoff by proposing a heterogenous visual search (HVS) system leading to 80 fold and 23 fold cost reduction for challenging retrieval problems on fashion (DeepFashion2) and face (IJB-C) images. This is achieved with marginal 0.3% and 1.6% drop in accuracy. Key to developing an HVS system is to ensure representational compatibility between the query and gallery embedding models.

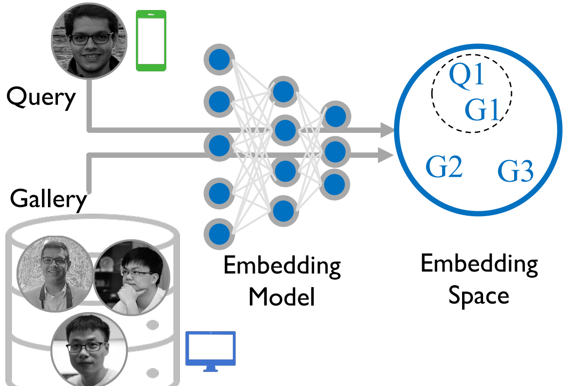

Fig. 1 A homogeneous visual search system with a shared embedding model.

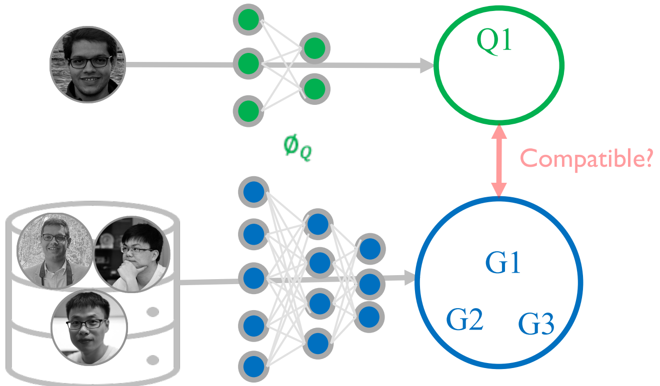

Fig. 2 A heterogeneous visual search system with decoupled embedding models.

Compatibility Criterion

Given models (ϕ_Q, ϕ_G)

Such that $Size(\phi_Q) << Size(\phi_G)$

$M(\phi_Q,\phi_Q) < M(\phi_Q,\phi_G) < M(\phi_G,\phi_G)$

M is a metric such as Top-k.

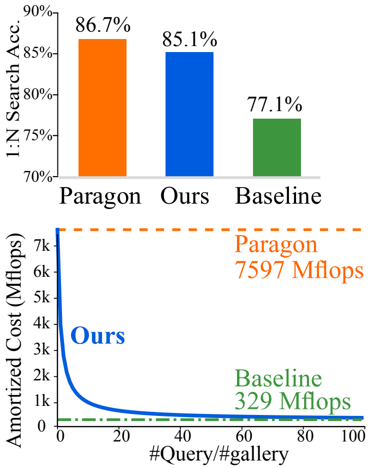

Fig. 3 Accuracy (top) and efficiency (bottom) of a heterogeneous visual search system.

Compatibility through weights

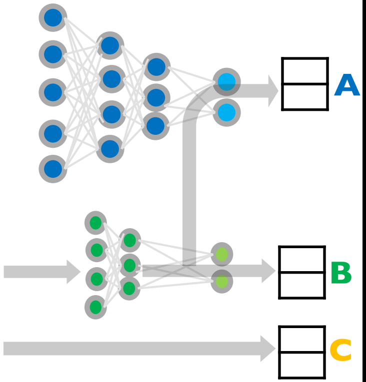

Fig. 4 Weight level compatibility.

Minimize BCT [1] Loss $\mathcal{L} = CE(\mathbf{B},\mathbf{C}) + CE(\mathbf{A},\mathbf{B})$

Compatibility through architecture (CMP-NAS)

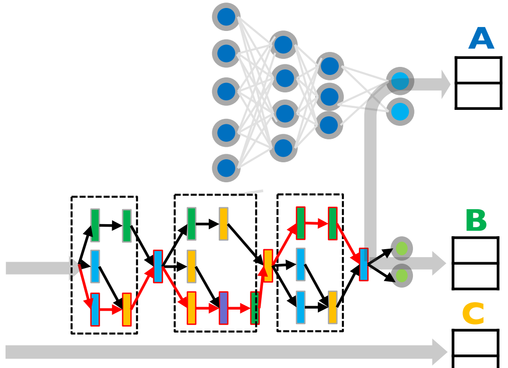

Fig. 5 Architecture level compatibility.

1. Sample architecture $\mathbf{a} \sim U(\Omega)$[2].
2. Minimize BCT [1] Loss $\mathcal{L} = CE(\mathbf{B},\mathbf{C}) + CE(\mathbf{A},\mathbf{B})$ for $\mathbf{a}$.
3. Maximize reward $a^* = argmax_a(\mathcal{R})$

Key Results

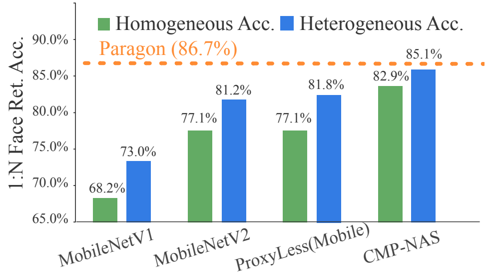

Fig. 6 Face Retrieval on IJB-C.

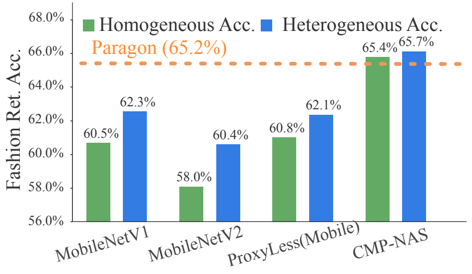

Fig. 7 Fashion Retrieval on DeepFashion-2.

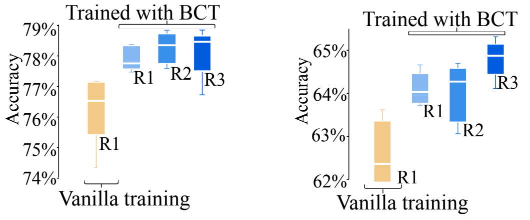

Fig. 8 Comparison with traditional NAS on the Faces (left) and Fashion (right) datasets.

[1] Shen, Yantao, et al. "Towards backward-compatible representation learning." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
[2] Guo, Zichao, et al. "Single path one-shot neural architecture search with uniform sampling." *European Conference on Computer Vision*. Springer, Cham, 2020.